ELSEVIER

Theoretical Computer Science 161 (1996) 93–108

Theoretical Computer Science

Flatwords and Post Correspondence Problem☆

Tero Harju, Marjo Lipponen *, Alexandru Mateescu

Academy of Finland and the Mathematics Department, University of Turku, 20500 Turku, Finland

Received February 1995; revised April 1995
Communicated by A. Salomaa

Abstract

We investigate properties of flatwords and k-flatwords. In particular, these words are studied in connection with Post Correspondence Problem (PCP). An open problem occurs: where is the borderline between the decidability and the undecidability of k-flat PCP over an alphabet with n symbols? Our main results concern the related new types of prime solutions of PCP.

1. Introduction

Let Σ be an alphabet. The set of all words over Σ is denoted by Σ^* and λ denotes the empty word. The length of a word w over Σ is denoted by $|w|$.

Consider the following version of the *Latin product*, denoted by $\diamond$. If $u = a_1a_2\dots a_n$ and $v = b_1b_2\dots b_m$ are words over Σ, then

$$u \diamond v = \begin{cases} a_1a_2\dots a_nb_2b_3\dots b_m & \text{if } a_n = b_1, \\ a_1a_2\dots a_nb_1b_2\dots b_m & \text{if } a_n \neq b_1. \end{cases}$$

By definition, $u \diamond \lambda = \lambda \diamond u = u$. It is easy to observe that the ordered system $\mathcal{M} = (\Sigma^*, \diamond, \lambda)$ is a monoid, called the *Latin monoid*. Let $Fl(\Sigma)$ be the submonoid of $\mathcal{M}$ generated by Σ. For instance, if $\Sigma = \{a, b\}$, then $Fl(\Sigma) = \{\lambda, a, b, ab, ba, aba, bab, \dots\}$, i.e. $Fl(\Sigma)$ consists of all words free of square letters. Such words will be referred as *flatwords*.

The Latin product, see [6, 8] is particularly important in enumeration problems in graph theory.

Flatwords occur in a natural way in problems concerning concurrent processes with re-entrant routines. A *re-entrant routine* may be executed concurrently by more than one process, see [1]. The trace of the parallel execution of two or more processes

☆ The work reported here has been supported by the Project 11281 of the Academy of Finland.
* Corresponding author. Email: marlip@utu.fi.

SSDI 0304-3975(95)00092-5

with re-entrant routines is always a set of flatwords. For more details concerning this problem, see [7].

In this paper we introduce k-flatwords and investigate some of their properties. We also study the relationship between k-flatwords and the Post Correspondence Problem. This is done in two ways: we define a new type of Post Correspondence Problem, k-flat PCP, and new types of primitive solutions for the instance $(g,h) = PCP$.

The suitability of the Post Correspondence Problem for reduction arguments is due to the fact that in some sense the essence of computations is captured by PCP. Thus simple solutions of PCP mean simplifications of computations and the results contribute on an abstract level to the understanding of computations.

Definition 1.1. Assume that $\Sigma = \{a_1, a_2, \ldots, a_n\}$ and let $k \geqslant 1$ be a fixed number. A word $w \in \Sigma^*$ is a *k-flatword* iff $|w| \leqslant 1$ or $w = x_1 x_2 \ldots x_m$ where $m \geqslant 2$ and $x_r \in \Sigma$, $1 \leqslant r \leqslant m$, and for any i_0 with $2 \leqslant i_0 \leqslant m$,

$$x_{i_0} \neq x_j$$

for all j such that $\max(1, i_0 - k) \leqslant j \leqslant i_0 - 1$.

Remark 1.1. Observe that a word w is a k-flatword if and only if all subwords of w of length at most $k+1$ have no multiple occurrences of letters.

Example 1.1. Let $\Sigma = \{a,b,c,d\}$ and the words w_1 and w_2 be the following:

$$w_1 = abcbcad, \qquad w_2 = cdabcad.$$

It is easy to observe that w_1 is a 1-flatword but not a 2-flatword because $x_4 = b = x_2$. On the other hand, the word w_2 is a 2-flatword but not a 3-flatword because $x_6 = a = x_3$.

Remark 1.2. Note that 1-flatwords are exactly the flatwords. Hence, a word w is a flatword iff $|w| \leqslant 1$ or $w = x_1 x_2 \ldots x_m$, where $x_j \in \Sigma$ and $x_i \neq x_{i-1}$ for $i = 2, \ldots, m$.

Notation. The set of all k-flatwords over an alphabet Σ is denoted by $Fl_k(\Sigma)$ and the set of all k-flatwords of length m by $Fl_k^{(m)}(\Sigma)$, i.e.,

$$Fl_k^{(m)}(\Sigma) = \{w \in Fl_k(\Sigma) \mid |w| = m\}.$$

Obviously, if $k \leqslant k'$, then

$$Fl_{k'}(\Sigma) \subseteq Fl_k(\Sigma).$$

Remark 1.3. If $\mathrm{card}(\Sigma) = n$ and $k \geqslant n$, then $Fl_k(\Sigma)$ is a finite set. For instance, if $\Sigma = \{a,b\}$, then

$$Fl_2(\Sigma) = \{\lambda, a, b, ab, ba\}.$$

Moreover, for any k, $k \geqslant n$, $Fl_k(\Sigma) = Fl_n(\Sigma)$.

2. Properties of k-flatwords

In this section we present some general properties of k-flatwords.

Proposition 2.1. *Let Σ be an alphabet such that* $\mathrm{card}(\Sigma) = n$. *Then*

$$\mathrm{card}(Fl_n(\Sigma)) = n! \sum_{j=0}^{n} \frac{1}{j!}.$$

Proof. For any number j, $1 \leqslant j \leqslant n$, the number of n-flatwords of length j can be counted as follows: There are n possibilities to choose the first symbol. For each, fixed, first symbol there are $n-1$ possibilities to choose the second symbol, etc. Hence, the number of n-flatwords of length j is

$$n(n-1)\cdots(n-j+1).$$

For $j > n$ there are no n-flatwords of length j because $\mathrm{card}(\Sigma) = n$. Notice that λ is also an n-flatword. Therefore,

$$\mathrm{card}(Fl_n(\Sigma)) = 1 + n + n(n-1) + \cdots + n(n-1)\cdots 2 \cdot 1 = n! \sum_{j=0}^{n} \frac{1}{j!}. \qquad \square$$

By Remark 1.3 we have now:

Corollary 2.1. *If* $\mathrm{card}(\Sigma) = n$, *then for any k, $k \geqslant n$,*

$$\mathrm{card}(Fl_n(\Sigma)) = n! \sum_{j=0}^{n} \frac{1}{j!}.$$

It is easy to observe that for $1 \leqslant k < n$ the set $Fl_k(\Sigma)$ is infinite.

Proposition 2.2. *Let Σ be an alphabet,* $\mathrm{card}(\Sigma) = n$, *and let k be a fixed number, $1 \leqslant k < n$. The number of k-flatwords of length p, $p \geqslant 0$, is given by the formula:*

$$\mathrm{card}(Fl_k^{(p)}(\Sigma)) = \begin{cases} \dfrac{n!}{(n-p)!} & \text{if } 0 \leqslant p \leqslant k, \\ \dfrac{n!}{(n-k)!}(n-k)^{p-k} & \text{if } p > k. \end{cases}$$

Proof. Assume that $0 \leqslant p \leqslant k$. The number of k-flatwords of length p is counted as in the proof of Proposition 2.1.

Let now $p > k$. In the same way as before the first k symbols can be chosen in $n(n-1)\cdots(n-k+1)$ ways. By Definition 1.1, there are $n-k$ ways to choose the next symbol. In fact, the same is true for all the remaining $p-k$ symbols because every time the new symbol is compared to the previous k symbols. Therefore, the number of k-flatwords of length p, $p > k$, is

$$\frac{n!}{(n-k)!}(n-k)^{p-k}. \qquad \square$$

Example 2.1. Let $\Sigma = \{a,b\}$ and $k = 1$. For any p, $p \geqslant 1$, there are exactly two flatwords of length p. If p is even, the flatwords of length p are $(ab)^{p/2}$ and $(ba)^{p/2}$; otherwise, the flatwords of length p are $(ab)^{(p-1)/2}a$ and $(ba)^{(p-1)/2}b$.

Proposition 2.3. *Let Σ be an alphabet and $k \geqslant 1$. The language $Fl_k(\Sigma)$ is a regular language.*

Proof. Consider the set

$$F = \{awa \mid a \in \Sigma, w \in \Sigma^* \text{ and } |w| \leqslant k-1\}.$$

It is clearly finite. Using Remark 1.1,

$$Fl_k(\Sigma) = \Sigma^* - \Sigma^* F \Sigma^*.$$

Therefore $Fl_k(\Sigma)$ is a regular language. $\square$

Comment. According to the previous proof, every $Fl_k(\Sigma)$ language can be defined by an extended star-free regular expression. From the Schützenberger's theorem, the syntactical monoid of each such language is aperiodic, i.e., the monoid has only trivial subgroups.

Moreover, for flatwords the set F has the property that $F \subseteq \Sigma^2$, and thus $Fl_1(\Sigma)$ is a local set, see [5].

Definition 2.1. A *k-flat language* is a language L such that $L \subseteq Fl_k(\Sigma)$.

Corollary 2.2. *Given $k \geqslant 1$, for every context-free language L the property of being a k-flat language is decidable. Given a context-free L, the smallest k such that L is k-flat is effectively computable.*

Proof. The claims follow from Proposition 2.3, Remark 1.3 and from the fact that it is decidable whether a context-free language is a subset of a regular language. $\square$

Comment. The family of k-flat languages is closed under union, intersection, left and right quotient, pref, suf and sub but it is not closed under catenation, Kleene star or complementation.

Proposition 2.4. *Let Σ be an alphabet,* $\text{card}(\Sigma) = k + 1$. *If L is a context-free language and $L \subseteq Fl_k(\Sigma)$, then L is a regular language.*

Proof. Observe that the set $\Sigma = \{a_1, a_2, \ldots, a_{k+1}\}$ has $(k+1)!$ ordered subsets with exactly $k+1$ distinct symbols. Denote these subsets by $s_1, s_2, \ldots, s_p$ where $p = (k+1)!$ and let $S = \{s_1, s_2, \ldots, s_p\}$. It is easy to see that for each word w, $w \in Fl_k(\Sigma)$, there exist a unique subset $s \in S$,

$$s = (a_{i_1}, a_{i_2}, \ldots, a_{i_{k+1}}),$$

a unique number j, $j \geqslant 0$, and a unique number r, $0 \leqslant r < k+1$, such that

$$w = (a_{i_1} a_{i_2} \ldots a_{i_{k+1}})^j a_{i_1} a_{i_2} \ldots a_{i_r}.$$

If $r = 0$, then $a_{i_1} a_{i_2} \ldots a_{i_r} = \lambda$.

The converse is obvious. Therefore, $Fl_k(\Sigma)$ has a finite decomposition of disjoint subsets:

$$Fl_k(\Sigma) = \bigcup_{s \in S} \bigcup_{r=0}^{k} F_{s,r},$$

where

$$F_{s,r} = \{(a_{i_1} a_{i_2} \ldots a_{i_{k+1}})^j a_{i_1} a_{i_2} \ldots a_{i_r} \mid j \geqslant 0\}.$$

Since $L \subseteq Fl_k(\Sigma)$, we have

$$L = L \cap Fl_k(\Sigma) = L \cap \left(\bigcup_{s \in S} \bigcup_{r=0}^{k} F_{s,r} \right) = \bigcup_{s \in S} \bigcup_{r=0}^{k} (L \cap F_{s,r}).$$

Denote the set $L \cap F_{s,r}$ briefly by $L_{s,r}$. Let x, y be two new symbols and define the morphism $h_{s,r} : \{x, y\}^* \to \Sigma^*$,

$$h_{s,r}(x) = a_{i_1} a_{i_2} \ldots a_{i_{k+1}}, \quad h_{s,r}(y) = a_{i_1} a_{i_2} \ldots a_{i_r}.$$

Then

$$h_{s,r}^{-1}(L_{s,r})/y = \{x^j \mid h_{s,r}(x^j y) \in L_{s,r}\}$$

is also a context-free language and since $h_{s,r}^{-1}(L_{s,r})/y \subseteq x^*$, it is a regular language. Consequently,

$$(h_{s,r}^{-1}(L_{s,r})/y) \cdot y = h_{s,r}^{-1}(L_{s,r})$$

is a regular language, too. Thus $L_{s,r}$ is a regular language for every $s \in S$ and $0 \leqslant r < k+1$. Finally, being a finite union of regular languages, L is a regular language. □

The previous result does not hold if L is a context-free language, but $L \subseteq Fl_{k-1}(\Sigma)$.

Example 2.2. Let $\Sigma = \{a, b, c\}$. The language

$$L = \{(ab)^n c (ab)^n \mid n \geqslant 0\}$$

is, clearly, context-free and $L \subseteq Fl_1(\Sigma)$. However, L is not a regular language. □

Definition 2.2. Let Σ be an alphabet and w a word over Σ, $w = x_1^{i_1}x_2^{i_2}\ldots x_n^{i_n}$, where $x_j \in \Sigma$, $i_j \geqslant 1$, $1 \leqslant j \leqslant n$ and $x_j \neq x_{j-1}$, $j = 2,3,\ldots,n$. The *flat image* of w, $fl(w)$, is defined by

$$fl(w) = x_1x_2\ldots x_n.$$

By definition, $fl(\lambda) = \lambda$.

Comment. Let φ be the inclusion function of Σ into the Latin monoid $\mathcal{M} = (\Sigma^*, \diamond, \lambda)$. Then the unique extension of φ to a morphism from Σ^* to $\mathcal{M}$ is exactly the fl function.

Moreover, a word w is a flatword if and only if $w = fl(w)$.

Let u, v be two words over Σ,

$$u = x_1^{i_1}x_2^{i_2}\ldots x_n^{i_n},$$

$$v = y_1^{j_1}y_2^{j_2}\ldots y_m^{j_m},$$

where $x_i, y_j \in \Sigma$ for all i,j. Define the relation $\leqslant_{fl}$ as follows:

$$u \leqslant_{fl} v \quad \text{iff} \quad fl(u) = fl(v) \quad \text{and} \quad (i_1, i_2, \ldots, i_n) \leqslant (j_1, j_2, \ldots, j_n),$$

where the inequality $\leqslant$ is the usual componentwise order between two vectors of positive integers. The condition $fl(u) = fl(v)$ implies, clearly, that $n = m$.

The following proposition shows that $\leqslant_{fl}$ is a partial order compatible with the monoid structure of Σ^*.

Proposition 2.5. *The relation $\leqslant_{fl}$ is a (partial) order between words. If $u \leqslant_{fl} v$, then for any word w, $uw \leqslant_{fl} vw$ and $wu \leqslant_{fl} wv$.*

3. PCP and k-flatwords

The Post Correspondence Problem (PCP) is one of the most common ways to prove undecidability in formal language theory. On the other hand, PCP provides a model of computation. The time and space complexity, the determinism and the nondeterminism can be expressed using PCP, see [2, 3, 17]. Therefore, relationships between PCP and k-flatwords might shed light also on the other concepts mentioned.

Let Σ and Δ be two alphabets, $\text{card}(\Sigma) = n$ and $\text{card}(\Delta) \geqslant 2$.

Assume that

$$g, h : \Sigma^* \longrightarrow \Delta^+$$

are two morphisms. The pair (g,h) is called an instance of *PCP*(n). The instance has a solution if there exists $w \in \Sigma^+$ such that $g(w) = h(w)$.

It is known that *PCP*(n) with $n \leqslant 2$ is a decidable problem, see [9], whereas *PCP*(n), with $n \geqslant 9$ is an undecidable problem, see [19, 18]. For the remaining n, $3 \leqslant n \leqslant 8$, the status of *PCP*($n$) is still open, see [10].

Definition 3.1. Let k,n be integers such that $k \geqslant 1$, $n \geqslant 1$. k-*flat PCP*(n) is the usual *PCP*(n) problem except that an instance (g,h) has a solution if and only if there exists a nonempty k-flatword w, such that $g(w) = h(w)$.

Proposition 3.1. *For any* $k, k \geqslant 1$, k-*flat PCP*(k+9) *is undecidable.*

Proof. By [4, 18] there is a family $F = \{(g_u, h_v) \mid u, v \in B^*\}$ of pairs of nonerasing morphisms such that *PCP* is undecidable for these. Here the alphabet for the morphisms is a 9-letter alphabet $A = C \cup \{d, e\}$, and the minimal solutions are of the form dwe, where w does not contain d or e. In fact, for $a \neq d$ the images $g_u(a)$ and $h_v(a)$ begin with a different letter.

Let $x_1, x_2, \ldots, x_{k+1}$ be $k+1$ new letters, and denote $u = x_1 x_2 \ldots x_{k+1}$. Define for each instance (g,h) from F a new pair (g', h') as follows.

Let $\alpha(a_1 a_2 \ldots a_n) = a_1 u a_2 u \ldots u a_n$ for each $a_1 a_2 \ldots a_n$, where a_i are letters, $i = 1, 2, \ldots, n$. Now, in the alphabet $A \cup \{f_1, f_2, \ldots, f_k\}$ define

$$
\begin{aligned}
g'(a) &= \alpha(g(a))x_1, & h'(a) &= x_{k+1}\alpha(h(a)) & &\text{for } a \in C,\\
g'(d) &= \alpha(g(d))x_1, & h'(d) &= \alpha(h(d)), & &\\
g'(e) &= \alpha(g(e)), & h'(e) &= x_{k+1}\alpha(h(e)), & &\\
g'(f_i) &= x_{i+1}, & h'(f_i) &= x_i, & &1 \leqslant i \leqslant k.
\end{aligned}
$$

Let $v = f_1 f_2 \ldots f_k$ and let $\alpha'(a_1 a_2 \ldots a_n) = a_1 v a_2 v \ldots v a_n$ for each $a_1 a_2 \ldots a_n$, where a_i are letters, $i = 1, 2, \ldots, n$.

If (g', h') has a solution w', then necessarily $w' = dva_1va_2 \ldots va_ne$ for some $w = da_1a_2 \ldots a_ne$, i.e., $w' = \alpha'(w)$. It is also easy to see that if w is a solution of (g,h), then $\alpha'(w)$ is a solution of (g', h').

Hence, k-flat *PCP*($k + 9$) is undecidable. □

In particular,

Corollary 3.1. *For any* $n, n \geqslant 10$, 1-*flat PCP*(n) *is undecidable.*

Proposition 3.2. *For any* k, $k \geqslant 1$, *there exists a number* n, *such that* k-*flat PCP*(n) *is a decidable problem.*

Proof. Define $n = k$. According to Proposition 2.1, the set $Fl_n(\Sigma)$ is finite and hence we need to verify only for a finite number of k-flatwords w whether $g(w) = h(w)$. □

The previous result can be extended for some cases where $Fl_n(\Sigma)$ is an infinite set.

Proposition 3.3. k-*flat PCP*($k + 1$) *is a decidable problem for any* k, $k \geqslant 1$.

Proof. Let $\Sigma = \{a_1, a_2, \ldots, a_{k+1}\}$. There are $(k+1)!$ ordered subsets of $(k+1)$ distinct elements. In the same way as in the proof of Proposition 2.4, for each word w, $w \in Fl_k(\Sigma)$, there exist a unique subset $s = (a_{i_1}, a_{i_2}, \ldots, a_{i_{k+1}})$, a unique number j, $j \geqslant 0$, and a unique number r, $0 \leqslant r < k+1$, such that

$$w = (a_{i_1} a_{i_2} \ldots a_{i_{k+1}})^j a_{i_1} \ldots a_{i_r}. \tag{1}$$

Let (g,h) be an instance of $PCP(k+1)$. We can now verify whether this instance has a solution of the form (1) in the following way:

Case I: Assume that $r = 0$. Then

$$g(w) = g((a_{i_1} a_{i_2} \ldots a_{i_{k+1}})^j) = (g(a_{i_1} a_{i_2} \ldots a_{i_{k+1}}))^j,$$
$$h(w) = (h(a_{i_1} a_{i_2} \ldots a_{i_{k+1}}))^j.$$

The equality $g(w) = h(w)$ is possible iff $g(a_{i_1} a_{i_2} \ldots a_{i_{k+1}}) = h(a_{i_1} a_{i_2} \ldots a_{i_{k+1}})$, which is obviously decidable.

Case II: Assume now that $r > 0$. Denote, briefly, $|g(a_{i_1} a_{i_2} \ldots a_{i_{k+1}})| = t_1$ and $|h(a_{i_1} a_{i_2} \ldots a_{i_{k+1}})| = t_2$.

Case IIa: If $t_1 = t_2$, then $g(w) = h(w)$ if and only if $g(a_{i_1} a_{i_2} \ldots a_{i_{k+1}}) = h(a_{i_1} a_{i_2} \ldots a_{i_{k+1}})$ and $g(a_{i_1} a_{i_2} \ldots a_{i_r}) = h(a_{i_1} a_{i_2} \ldots a_{i_r})$ where the last two equalities are decidable.

Case IIb: Assume now that $t_1 \neq t_2$ and denote $|g(a_{i_1} a_{i_2} \ldots a_{i_r})| = q_1$ and $|h(a_{i_1} a_{i_2} \ldots a_{i_r})| = q_2$. The equation $g(w) = h(w)$ implies that $|g(w)| = |h(w)|$, i.e.,

$$jt_1 + q_1 = jt_2 + q_2.$$

Hence, $j(t_1 - t_2) = q_2 - q_1$ and therefore, j has a unique value, $j = (q_2 - q_1)/(t_1 - t_2)$. If j is a positive integer we need to verify the equality $g(w) = h(w)$ only for this particular j. Repeat this procedure for each ordered subset s. Again this is decidable. □

Corollary 3.2. *For all n, $n \geqslant 10$, 1-flat PCP(n) is undecidable and $(n-1)$-flat PCP(n) is decidable.*

Open Problem. Assume that $n \geqslant 10$ and let k be a fixed number, $1 < k < (n-1)$. Is k-flat $PCP(n)$ decidable or not? Where is the borderline between decidability and undecidability in this case?

4. Flat prime solutions of the Post Correspondence Problem

We now return to the usual Post Correspondence Problem. The study of prime solutions of *PCP* was initiated in [20] and later extended in [11–16].

A prime solution of an instance $(g,h) = PCP$ is a solution that is somehow simpler than the other solutions. More specifically, we fix an order "$\leqslant$" between words. (The order can be defined in such a way that it satisfies some practical purposes.) If $u \leqslant v$, we say that u is simpler than v. A prime solution of $(g,h) = PCP$ is now a solution which is minimal with respect to this order. Explicit cases of this general idea will be discussed below.

Let g and h be nonerasing morphisms of Σ^* into Δ^*, where Σ and Δ are finite alphabets. The *equality set* between g and h is defined by

$$E(g,h) = \{w \in \Sigma^+ \mid g(w) = h(w)\}.$$

For a word w over Σ^*, we now consider the sets of words obtained from w by removing a final subword, a subword or a scattered subword, respectively. Define

$$\begin{aligned} \text{fin}\,(w) &= \{v \mid w = vx, \text{ for some } x \in \Sigma^*\}, \\ \text{sub}\,(w) &= \{v_1v_2 \mid w = v_1xv_2, \text{ for some } v_1,v_2,x \in \Sigma^*\}, \\ \text{scatsub}\,(w) &= \{v_1 \cdots v_k \mid w = x_1v_1 \cdots x_kv_kx_{k+1}, \text{ for some } x_i,v_i \in \Sigma^*\}. \end{aligned}$$

We can now determine three further sets, as follows:

$$\begin{aligned} F(g,h) &= \{w \in E(g,h) \mid \text{fin}\,(w) \cap E(g,h) = \{w\}\}, \\ S(g,h) &= \{w \in E(g,h) \mid \text{sub}\,(w) \cap E(g,h) = \{w\}\}, \\ P(g,h) &= \{w \in E(g,h) \mid \text{scatsub}\,(w) \cap E(g,h) = \{w\}\}. \end{aligned}$$

Words in the three sets are called *F-prime*, *S-prime* and *P-prime solutions* for the instance $PCP = (g,h)$, respectively.

It is a direct consequence of the definitions that

$$P(g,h) \subseteq S(g,h) \subseteq F(g,h) \subseteq E(g,h).$$

One or both of the first inclusions may be strict whereas the third inclusion is always strict, provided $E(g,h)$ is nonempty. If $E(g,h)$ is nonempty then so must be the three other sets. In addition, each of the four sets is recursive.

The triple (p,s,f), where p, s and f are the cardinalities of the sets $P(g,h)$, $S(g,h)$ and $F(g,h)$, respectively, is defined to be the *primality type* of the instance $PCP = (g,h)$. Thus p, s and f are nonnegative integers or ∞. These triples were fully characterized in [20].

Proposition 4.1. *A triple* (p,s,f) *is a primality type if and only if either* $p = s = f = 0$, *or else each of the following conditions* (i)–(iii) *holds*: (i) $1 \leqslant p \leqslant s \leqslant f$, (ii) *p is finite*, (iii) *if* $s < f$ *then* $f = \infty$. *An example for each possible type can be effectively constructed.*

We now consider words that can describe a prime solution, for some instance of PCP. We say that a word w over Σ is a *P-word* if, for some instance (g,h), w is in $P(g,h)$. *S-words* and *F-words* are defined similarly.

The following result established in [15, 13] gives a way of constructing words that are not F-words (and thus not S- and P-words either). Here $\psi(w)$ means the Parikh vector.

Proposition 4.2. *A word w is an F-word iff w has no nontrivial prefix w_1 satisfying $\psi(w_1) = r\psi(w)$ for some (rational) number r.*

The proposition gives, among others, the following examples: $ab^2a^2b^4$, $ab^6a^3b^2$.

Denote the set of P-words by P, S-words by S and F-words by F. In [12] the following result was proved.

Proposition 4.3. *$P \subset S \subset F$ where all the inclusions are proper if* $\text{card}(\Sigma) \geqslant 3$.

For instance, in a three-letter alphabet the word $(ab)^2(abc)^3$ is an F-word but not an S-word and $a^2b^3c^3a^4b^3c^3a^3$ is an S-word but not a P-word.

We now extend the definition of prime solutions to concern k-flatwords.

Definition 4.1. A solution $\alpha \in E(g,h)$ is $F_{k\text{-flat}}$-prime iff no k-flat suffix of α, that is, a suffix which is a k-flatword, can be removed and still obtain a solution. Analogously, $S_{k\text{-flat}}$-prime and $P_{k\text{-flat}}$-prime solutions correspond to removing a k-flat subword and a k-flat scattered subword, respectively.

Observe that an $F_{k\text{-flat}}$-prime solution need not be a k-flatword.

The sets of 1-flat prime solutions are denoted briefly by $F_{\text{flat}}(g,h)$, $S_{\text{flat}}(g,h)$ and $P_{\text{flat}}(g,h)$. By definition,

$$P_{k\text{-flat}}(g,h) \subseteq S_{k\text{-flat}}(g,h) \subseteq F_{k\text{-flat}}(g,h).$$

Theorem 4.1. *In every alphabet Σ,* $\text{card}(\Sigma) = n$, *we have the following hierarchy*:

$$\begin{array}{ccccccc}
P_{n\text{-flat}}(g,h) & \rightarrow & S_{n\text{-flat}}(g,h) & \rightarrow & F_{n\text{-flat}}(g,h) & \rightarrow & E(g,h), \\
\uparrow & & \uparrow & & \uparrow & & \\
\vdots & & \vdots & & \vdots & & \\
\uparrow & & \uparrow & & \uparrow & & \\
P_{2\text{-flat}}(g,h) & \rightarrow & S_{2\text{-flat}}(g,h) & \rightarrow & F_{2\text{-flat}}(g,h), & & \\
\uparrow & & \uparrow & & \uparrow & & \\
P_{\text{flat}}(g,h) & \rightarrow & S_{\text{flat}}(g,h) & \rightarrow & F_{\text{flat}}(g,h), & & \\
\uparrow & & \uparrow & & \uparrow & & \\
P(g,h) & \rightarrow & S(g,h) & \rightarrow & F(g,h). & &
\end{array}$$

Proof. By Definition 1.1, k-flatwords are also $(k-1)$-flatwords. But this means that if no $(k-1)$-flat suffix (respectively subword, scattered subword) can be removed this is true also for k-flat suffixes (resp. subwords, scattered subwords). Therefore,

$$F_{(k-1)\text{-flat}}(g,h) \subset F_{k\text{-flat}}(g,h),$$
$$S_{(k-1)\text{-flat}}(g,h) \subset S_{k\text{-flat}}(g,h),$$
$$P_{(k-1)\text{-flat}}(g,h) \subset P_{k\text{-flat}}(g,h).$$

In the same way,

$$F(g,h) \subset F_{\text{flat}}(g,h), \quad S(g,h) \subset S_{\text{flat}}(g,h), \quad P(g,h) \subset P_{\text{flat}}(g,h).$$

Thus, the sets form an increasing hierarchy which is, however, finite because k-flatwords equal n-flatwords if $k \geqslant n$ (see Remark 1.3). $\square$

Consider now the hierarchy of words. We prove that it is the same as in Theorem 4.1 except that it is infinite, all the inclusions are proper (which is by Proposition 4.3 already known for P, S and F) and that no other inclusions exist, that is, $P_{k\text{-flat}}$ and $S_{(k-1)\text{-flat}}$ as well as $S_{k\text{-flat}}$ and $F_{(k-1)\text{-flat}}$ are incomparable.

Consider first some examples and an interesting property of $F_{k\text{-flat}}$-words which is not true for F-words. The following proof shows also that there exists words that are not $F_{k\text{-flat}}$-words for any k.

Theorem 4.2. *The property of being an $F_{k\text{-flat}}$-word is not preserved by mirror image.*

Proof. The word 112212 cannot be an $F_{k\text{-flat}}$-word because $112212 \in E(g,h)$ implies that also $1122 \in E(g,h)$ and 12 is a k-flatword. On the other hand, $212211 \in F_{k\text{-flat}}(g,h)$, as shown by the instance

	1	2
g	a^2	a
h	a	a^2

$\square$

Example 4.1. Let the instance (g,h) be defined by

	1	2
g	ab	c
h	a	bc

Clearly, $E(g,h) = (12)^+$. Consider the word $(12)^n$, $n \geqslant 2$. It cannot be an $F_{k\text{-flat}}$-solution because 12 is a k-flatword and $(12)^{n-1}$ is a solution. Thus, none of the words $(12)^n$, $n \geqslant 2$, is $F_{k\text{-flat}}$-prime and

$$F_{k\text{-flat}}(g,h) = S_{k\text{-flat}}(g,h) = P_{k\text{-flat}}(g,h) = F(g,h) = S(g,h) = P(g,h) = \{12\}.$$

Example 4.2. Consider the instance (g,h) defined by

	1	2
g	a^2b	a^2
h	a	$(ba^2)^2$

Now $E(g,h) = (112)^+$. A word $(112)^n$ is an $F_{k\text{-flat}}$-solution for all n because no k-flat suffix can be removed. On the other hand,

$$(112)^n = 1\underbrace{121}12(112)^{n-2}$$

is not S_{flat}-prime because 121 is a flatword (but not a k-flatword, $k \geqslant 2$) and $(112)^{n-1}$ is a solution. Thus, for $k \geqslant 2$,

$$F_{k\text{-flat}}(g,h) = S_{k\text{-flat}}(g,h) = P_{k\text{-flat}}(g,h) = F_{\text{flat}}(g,h) = (112)^+,$$
$$S_{\text{flat}}(g,h) = P_{\text{flat}}(g,h) = F(g,h) = S(g,h) = P(g,h) = \{112\}.$$

Example 4.3. Consider now

	1	2
g	a^3b	a^3
h	a	$(ba^3)^3$

This time the words $(1112)^n$ are all P_{flat}-primes because none of the flat scattered subwords can be removed. Therefore,

$$P_{k\text{-flat}}(g,h) = S_{k\text{-flat}}(g,h) = F_{k\text{-flat}}(g,h) = (1112)^+,$$
$$P(g,h) = S(g,h) = F(g,h) = \{1112\}.$$

Theorem 4.3. *The word w^n, $n \geqslant 2$, is $F_{k\text{-flat}}$-prime iff w is not a k-flatword and w is an $F_{k\text{-flat}}$-prime.*

Proof. Clearly, w^n is a solution iff w is a solution. Thus, if a k-flat suffix of w can be removed, it can be done for w^n too. On the other hand, if w^n has a k-flat suffix u and $|u| > |w|$ then w is also a k-flatword. □

Theorem 4.4. *If $w = a_1 \ldots a_k a_{k+1}^2$, where $a_i \in \Sigma$ and $a_1 \ldots a_{k+1}$ is a k-flatword, then w^n is not $S_{k\text{-flat}}$-prime, for $n \geqslant 2$.*

Proof. The technique is similar as in Example 4.2. Consider

$$w^2 = a_1 \ldots a_k a_{k+1} \underbrace{a_{k+1} a_1 \ldots a_k a_{k+1}}_{\text{can be removed}} a_{k+1}. \qquad \square$$

Theorems 4.3 and 4.4 imply that there exists an $F_{k\text{-flat}}$-word that is not an $S_{k\text{-flat}}$-word. The same is true for $S_{k\text{-flat}}$-words and $P_{k\text{-flat}}$-words. For this we need the following result from [13]. A word w is *periodicity forcing* if $w \in E(g,h)$ implies that g and h are periodic, that is there exists a word u over Δ such that $g(i), h(i) \in u^+$ for each $i \in \Sigma$.

Proposition 4.4. *Let $\Sigma = \{a, b_2, \ldots, b_n\}$, $n \geqslant 2$ and $i_1, \ldots, i_n$, $j_2, \ldots j_n$ be such that $i_1 + i_n \neq 0$, $j_k \neq 0$ for $k = 2, \ldots n$ and $i_l \neq 0$ for $l = 2, \ldots n-1$. Then the word $w = avava$ where*

$$v = a^{i_1} b_2^{j_2} a^{i_2} b_3^{j_3} \ldots a^{i_{n-1}} b_n^{j_n} a^{i_n} \qquad (2)$$

is periodicity forcing.

Theorem 4.5. *For each k there exists an $S_{k\text{-flat}}$-word that is not a $P_{k\text{-flat}}$-word.*

Proof. Let n be the smallest number divisible by 3 and greater than k. Consider the following word $w = avava$ of the form (2). Let $i_l = 1$ for $l = 1,\ldots,n$ and $j_s = 2r+1$, for $s = 2,\ldots,n$ where $n = 3r$. Then

$$\psi(w) = (6r+3, 2(2r+1),\ldots,2(2r+1)) = (2r+1)(3,2,\ldots,2).$$

By Proposition 4.4 there exists a scattered subword $u = ab_2b_3\ldots b_nab_2b_3\ldots b_na$ which is an $(n-1)$-flatword and thus a k-flatword; hence, w is not a $P_{k\text{-flat}}$-solution. It is, however, an $S_{k\text{-flat}}$-solution because it is an S-solution by [13]. □

Example 4.4. Let $n=3$. Then $r=1$ and $j_l=2r+1=3$. Consider the word $w = avava$ where $v = ab_2^3ab_3^3a$, i.e.,

$$w = a^2b_2^3ab_3^3a^3b_2^3ab_3^3a^2.$$

Since $u = ab_2b_3ab_2b_3a$ is a 2-flatword and $ab_2^2ab_3^2a^2b_2^2ab_3^2a$ is a solution, w cannot be $P_{2\text{-flat}}$-prime or P_{flat}-prime.

Comment. Notice that in the previous example u is not a 3-flatword. This is why we have to choose n greater than k in the proof of Theorem 4.5.

Theorem 4.6. $P_{k\text{-flat}} \subset S_{k\text{-flat}} \subset F_{k\text{-flat}}$.

Theorem 4.7. *In every alphabet Σ, $\text{card}(\Sigma) = n$, there exists a word w that is a $P_{n\text{-flat}}$-word but not an $F_{(n-1)\text{-flat}}$-word.*

Proof. Let $\Sigma = \{a_1,\ldots,a_n\}$. The word

$$w = (a_1a_2\ldots a_na_1)^2$$

is not an $F_{(n-1)\text{-flat}}$-solution because $a_1a_2\ldots a_na_1$ is an $(n-1)$-flatword; nevertheless, since no n-flat scattered subwords can be removed, w is a $P_{n\text{-flat}}$-solution for the following instance:

	a_1	a_2	a_3	a_4	$\ldots$	a_{n-1}	a_n
g	$b_1b_2b_1$	b_3	b_4	b_5	$\ldots$	b_n	b_{n+1}
h	b_1	b_2b_1	b_3	b_4	$\ldots$	b_{n-1}	$b_nb_{n+1}b_1b_2$

where $\Delta = \{b_1,\ldots,b_{n+1}\}$. □

Theorem 4.8. *The sets $P_{k\text{-flat}}$ and $S_{(k-1)\text{-flat}}$ are incomparable.*

Proof. By Theorem 4.7 there are words in $P_{k\text{-flat}}$ that are not in $S_{(k-1)\text{-flat}}$. On the other hand, the word w in the proof of Theorem 4.5 is an S-solution and thus an $S_{k\text{-flat}}$-solution for all k but not a $P_{k\text{-flat}}$-word. □

Theorem 4.9. *The sets $S_{k\text{-flat}}$ and $F_{(k-1)\text{-flat}}$ are incomparable.*

Proof. By the proof of Theorem 4.7 there are words in $S_{k\text{-flat}}$ that are not in $F_{(k-1)\text{-flat}}$. Consider now the word w in Theorem 4.4,

$$w = (a_1 \dots a_k a_{k+1}^2)^2.$$

It is not an $S_{k\text{-flat}}$-word but it is an $F_{k\text{-flat}}$-solution for the instance

	a_1	a_2	a_3	$\dots$	a_{n-1}	a_n	a_{n+1}
g	b_1b_2	b_3	b_4	$\dots$	b_n	b_{n+1}^2	b_{n+1}
h	b_1	b_2	b_3	$\dots$	b_{n-1}	b_n	b_{n+1}^2

where $\Delta = \{b_1, \dots, b_{n+1}\}$.

The relation between F-words and S_{flat}-words still remains to be settled. By Proposition 4.2 we know that no word of the form w^i, $i \geqslant 2$ is an F-word. On the other hand, in [11] we proved that the word $w = 112122$ is periodicity forcing. Therefore, if $112122 \in E(g,h)$ also $12 \in E(g,h)$; hence, w is not an S_{flat}-word. It is, however, an F-solution for the instance (g,h) presented in the proof of Theorem 4.2. □

The following result shows that our hierarchy may collapse.

Theorem 4.10. *If $|F_{k\text{-flat}}(g,h)| < \infty$ then $F_{k\text{-flat}}(g,h) = S_{k\text{-flat}}(g,h) = F(g,h) = S(g,h)$.*

Proof. If $|F_{k\text{-flat}}(g,h)| < \infty$ then $f < \infty$ and $s = f = n$ by Theorem 4.1 and Proposition 4.1. This implies that $E(g,h) = \{w_1, \dots, w_n\}^+$. By Theorem 4.3 each word w_i must be a k-flatword; hence, $|F_{k\text{-flat}}(g,h)| = n$ and, moreover, $F_{k\text{-flat}}(g,h) = S_{k\text{-flat}}(g,h)$. □

Theorem 4.11. *If $f < \infty$ then $S_{k\text{-flat}}(g,h) = F(g,h) = S(g,h)$, or else $|S_{k\text{-flat}}(g,h)| = \infty$.*

Proof. In the same way as in the previous proof, $f < \infty$ implies that $s = f = n$ and $E(g,h) = \{w_1, \dots, w_n\}^+$. Assume now that $|S_{k\text{-flat}}(g,h)| < \infty$. If there exists an $S_{k\text{-flat}}$-solution that is not an F-solution, it must be of the form w_iw_j or w^i for some $w_i, w_j, w \in E(g,h)$. Consider first the case w_iw_j. If w_iw_j is an $S_{k\text{-flat}}$-solution but $(w_iw_j)^2$ is not, then $(w_iw_j)^2 = v_1xv_2$ where v_1v_2 is a solution and x is a k-flatword.

By Theorem 4.3 the words w_i, w_j cannot be k-flatwords; hence, $w_iw_j \in \text{sub}(v_1v_2)$. But this means that v_1v_2 cannot be a solution w, $w \in \{w_1, \dots, w_n\}$ or a catenation of solutions w since in both cases either w_i is a subword of w or vice versa, which is not possible since $S(g,h) = F(g,h)$.

The same is true for $S_{k\text{-flat}}$-solutions of the type w^i. Therefore, if there exists an $S_{k\text{-flat}}$-solution that is not an F-solution, then $|S_{k\text{-flat}}(g,h)| = \infty$. □

Theorem 4.12. *Let Σ be an alphabet,* card$(\Sigma) = n$*. There are no* $(n-1)$*-flatwords* w, $|w| \leqslant n$, *in the equality set of the instance* (g,h) *iff*

$$F_{k\text{-flat}}(g,h) = E(g,h) \quad \text{for } k \geqslant n.$$

Proof. By Theorem 2.1, the only k-flatwords, $k \geqslant n$, over Σ are $(n-1)$-flatwords. For instance, in the binary alphabet the only k-flatwords, $k \geqslant 2$, are

$$w = 1,\ 2,\ 12,\ 21.$$

If none of these is a solution then no solution u can be divided in such way that the suffix is a k-flatword and the prefix is a solution.

On the other hand, if there is such a solution w then by Theorem 4.3 the word w^n, $n \geqslant 2$, cannot be an $F_{k\text{-flat}}$-solution; hence, $F_{k\text{-flat}}(g,h) \neq E(g,h)$. □

Corollary 4.1. *If* $P_{n\text{-flat}}(g,h) = E(g,h)$ *(resp.* $S_{n\text{-flat}}(g,h) = E(g,h)$*) then there are no* $(n-1)$*-flatwords* w, $|w| \leqslant n$, *in* $E(g,h)$.

By Example 4.2 and Theorem 4.4 the converse is not true.

Theorem 4.13. *The property of being an* $F_{k\text{-flat}}$*-word is decidable.*

Proof. Let $w \in \Sigma^+$. The following algorithm shows whether w is an $F_{k\text{-flat}}$-word or not.

Step I: Find the shortest word u for which $w = u^n$.

Step II: If $n > 1$ and u is a k-flatword then w is not an $F_{k\text{-flat}}$-word by Theorem 4.3.

Step III: If u is an F-word (as in Proposition 4.2) then w is an $F_{k\text{-flat}}$-word by Theorem 4.3.

Step IV: Check for each division $u = u_1u_2$, where $\psi(u_1) = r\psi(u)$ for some $r < 1$, whether u_2 is a k-flatword. If this is not true for any u_2 then w is an $F_{k\text{-flat}}$-word; otherwise, w is not an $F_{k\text{-flat}}$-word.

The morphisms g and h for which $w \in F_{k\text{-flat}}(g,h)$ in steps III and IV can be effectively computed as shown in [13]. □

We have thus characterized the set of $F_{k\text{-flat}}$-words. An important open problem, both in decidability theory and in combinatorics on words, is to find characterizations for $P_{k\text{-flat}}$- and $S_{k\text{-flat}}$-words as well. At the moment, even the recursiveness of these sets remains open.

References

[1] T. Axford, *Concurrent Programming Fundamental Techniques for Real-Time and Parallel Software Design* (Wiley, New York, 1989).

[2] R.V. Book and F.J. Brandenburg, Representing complexity classes by equality sets, Lecture Notes in Computer Science, Vol. 71 (Springer, Berlin, 1979), 49–57.

[3] R.V. Book and F.J. Brandenburg, Equality sets and complexity classes, *SIAM J. Comput.* **9** (1980) 729–743.
[4] V. Claus, Some remarks on PCP(k) and related problems, *EATCS Bull.*, **12** (1980) 54–61.
[5] S. Eilenberg, *Automata, Languages, and Machines, Vol. A* (Academic Press, New York, 1974).
[6] J.S. Golan, *The Theory of Semirings with Applications in Mathematics and Theoretical Computer Science* (Longman, New York 1992).
[7] J.S. Golan, A. Mateescu and D. Vaida, Towards a unified theory of sequential, parallel and semi-parallel processes, submitted.
[8] M. Gondram and M. Minoux, *Graphs and Algorithms* (Wiley-interscience, New York 1984).
[9] A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, The (generalized) Post Correspondence Problem with lists consisting of two words is decidable, *Theoret. Comput. Sci.* **21** (1982) 119–144.
[10] J. Karhumäki, On recent trends in formal language theory, in: T. Ottmann, ed., *Proc. 14th ICALP, 1987*, Lecture Notes in Computer Science, Vol. 267 (Springer, Berlin, 1987) 138–162.
[11] M. Lipponen, Primitive Words and Languages Associated to PCP, *EATCS Bull.* **53** (1994) 217–226.
[12] M. Lipponen, Post Correspondence Problem: words possible as primitive solutions, in: *Proc. ICALP-95*, (Springer, Berlin), to appear.
[13] M. Lipponen, On *F*-prime solutions of the Post Correspondence Problem, in: *Proc. 2nd Internat. Conf. on Developments in Language Theory*, Magdeburg, 1995, to appear.
[14] M. Lipponen and Gh. Păun, Strongly prime PCP words, *Discrete Appl. Math.*, to appear.
[15] A. Mateescu and A. Salomaa, PCP-prime words and primality types, *RAIRO Inform. Théorique* **27** (1993) 57–70.
[16] A. Mateescu and A. Salomaa, On simplest possible solutions for Post Correspondence Problems, *Acta Inform.* **30** (1993) 441–457.
[17] A. Mateescu, A. Salomaa, K. Salomaa and S. Yu, P, NP and Post Correspondence Problem, *Inform. and Comput.*, to appear.
[18] J.J. Pansiot, A note on Post's Correspondence Problem, *Inform. Proc. Lett.* **12** (1981) 233.
[19] E. Post, A variant of a recursively unsolvable problem, *Bull. Amer. Math. Soc.* **53** (1946) 264–268.
[20] A. Salomaa, K. Salomaa and S. Yu, Primality types of instances of the Post Correspondence Problem, *EATCS Bull.* **44** (1991) 226–241.